Prototyping Fusion Center Information Sharing: Implementing Policy Reasoning Over Cross-Jurisdictional Data Transactions Occurring in a Decentralized Environment

K. Krasnow Waterman and Samuel Wang

ABSTRACT

In 2004, the White House and then Congress determined there should be an "Information Sharing Environment" that facilitates the flow of critical information for counterterrorism, related law enforcement, and disaster management activities. That work has been progressing but a major challenge is how to create technologies that: ensure compliance with laws and policies of the federal government, fifty states, and individual agencies; convey appropriate data that would support access control and privilege decisions in different jurisdictions; and achieve accountability and transparency for this activity. We have built a prototype of Fusion Center information sharing that shows significant progress in the representation of law in a policy language, the reasoning of that law over data transactions occurring in a web environment (internet or intranet), acquiring necessary information from authoritative sources wherever they reside in the decentralized environment, and providing both a binary response suitable for automated workflow implementation and a detailed justification suitable for human validation of the conclusion. In this paper, we briefly describe the technologies employed for serializing the data and policy, reasoning over the rules contained in the policy, and displaying the results to users. These combine to provide a powerful tool supporting a range of necessary governmental functions including access control, privilege management, audit, periodic reporting, and risk modeling.

INTRODUCTION

After 9/11, a cry arose within the United States that the terrorist attack could have been averted if government agencies had shared what they knew with each other. While the accuracy of that claim remains in debate, there is significant evidence that agencies were sharing less than expected and that they would operate more effectively if they shared more information. Three years later, having not made significant progress towards that goal, the White House issued an Executive Order mandating the creation of an Information Sharing Environment; this goal was reinforced by Congress later the same year when it was mandated in a new statute.1

In the years since the goal was set, an impediment to implementation has been identified. The sharing is mandated to be performed "[t]o the maximum extent consistent with applicable law." However, a gap exists between the laws and policies enacted by government to regulate the handling of information and the ability to enforce those policies in computer systems. There is a strong need to bridge that gap as more data is or is desired to be collected, shared, and manipulated. Responsible managers and interested citizens alike are seeking the means to ensure that systems more effectively implement rules about privacy, security, and the appropriate conduct of government business. But, while people can express rules with complex reasoning, context, and reference to information not contained in the subject data, information systems historically have not been able to process policies written this way.

For example, consider the following snippet of legislation enacted by the state of Maryland:

> A. Subject to the provisions of Regulation .12B, the Central Repository and other criminal justice agencies shall disseminate CHRI, be it conviction or nonconviction criminal history record information, to a criminal justice agency upon a request made in accordance with applicable regulations adopted by the Secretary. A criminal justice agency may

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
MAR 2011		**00-00-2011 to 00-00-2011**

4. TITLE AND SUBTITLE	5a. CONTRACT NUMBER
Prototyping Fusion Center Information Sharing:Implementing Policy Reasoning Over Cross-Jurisdictional Data Transactions Occurring In A Decentralized Environment	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **MIT,Cambridge,MA,02193**	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES
HOMELAND SECURITY AFFAIRS, SUPPLEMENT 3 (MARCH 2011),Government or Federal Purpose Rights License

14. ABSTRACT
In 2004, the White House and then Congress determined there should be an ?Information Sharing Environment? that facilitates the flow of critical information for counterterrorism, related law enforcement, and disaster management activities. That work has been progressing but a major challenge is how to create technologies that: ensure compliance with laws and policies of the federal government, fifty states, and individual agencies; convey appropriate data that would support access control and privilege decisions in different jurisdictions; and achieve accountability and transparency for this activity.We have built a prototype of Fusion Center information sharing that shows significant progress in the representation of law in a policy language, the reasoning of that law over data transactions occurring in a web environment (internet or intranet), acquiring necessary information from authoritative sources wherever they reside in the decentralized environment, and providing both a binary response suitable for automated workflow implementation and a detailed justification suitable for human validation of the conclusion. In this paper, we briefly describe the technologies employed for serializing the data and policy, reasoning over the rules contained in the policy, and displaying the results to users. These combine to provide a powerful tool supporting a range of necessary governmental functions including access control,privilege management, audit, periodic reporting,and risk modeling.

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**	**Same as Report (SAR)**	**11**	

> request this information from the Central Repository or another criminal justice agency only if it has a need for the information:
>
> (1) In the performance of its function as a criminal justice agency; or
>
> (2) For the purpose of hiring or retaining its own employees and agents.[2]

It is clear that the intent of this legislation is to regulate the transmission of sensitive criminal history record information so that it is only used for appropriate purposes. However, the interactions between this specific policy and other policies at the organization, state, and federal level could potentially be very complex, and it is not feasible for humans to reason over all of them simultaneously. In addition, the rules and terms used in policies often reference other policies and pieces of information located in different databases or organizations, which makes it difficult to efficiently verify compliance by hand. Finally, given the number of transactions that happen per day, if a violation does occur, it is difficult to verify exactly which information sharing transaction was non-compliant with the applicable policies.

Given that computers are already ubiquitous in data sharing environments due to the ease of sharing and aggregating information, it is worthwhile to investigate whether or not they can also solve the problems listed above. We built a prototype of an "accountable system" to address this challenge by using semantic web technology. Semantic web technology generally seeks to express data on the internet in a way such that machines can reason over the semantics of the data more readily. An accountable system is one that both knows which policies apply to which data (policy awareness), and one that can reason over complex policy and the details of data transactions. These two functions together allow organizations to fulfill their obligations in a transparent and policy-aware manner. In this project, we modeled transactions between Fusion Centers, locations where state and federal agencies work cooperatively to address terrorism, crime, and emergency response. Our prototype shows that the authoritative sources of information needed to make policy-based decisions can remain and be accessed wherever they reside in the decentralized environment.

This paper presents a prototype system that models the data sharing workflow in a Fusion Center environment, with the following features:

1. An effective way to represent real legislation and policies in a computer-readable language that can be reasoned upon.
2. A model where existing data can remain in disparate databases and servers which the reasoner can access on the fly during reasoning.
3. A reasoner which can analyze transactions with rules and then present a justification of why the transaction is or is not compliant.
4. A user interface which analysts and law enforcement can use to determine whether transactions are compliant with the applicable policies. The user interface is designed for end users who have neither a legal nor a technical background, presenting justifications in natural language to users.

The prototype demonstrates that such a reasoning system can be used to increase the amount of transparency and accountability in real data-sharing environments. Given any data sharing event, the reasoner can produce a transcript that shows exactly which pieces of data went into the decision, which parts of the law are relevant, and the apparent compliance or non-compliance with those parts.

BACKGROUND

Semantic Web and Linked Data

The primary motivation of the semantic web is that by associating metadata with data on the web, it enables computers to do more valuable computations than if computers did not know about the semantics of the data at hand. In particular, websites today are designed primarily for user consumption, in that machines have a hard time understanding the semantic content on any given page. If the pages also provide machine-readable metadata, automated

agents can more easily perform tasks on behalf of the user.

Linked Data is the notion that by associating a unique identifier (a URI) with each piece of data in question, it is possible to create unambiguous references between pieces of data. This ability to create relationships between disparate datasets greatly increases the utility of the data, and allows computers to reason over the relationships between data. In addition, it's no longer necessary to warehouse data in one centralized location, as data in one database can refer to data in another database by URI just as easily as it can refer to data in the same database. More details about Linked Data can be found in the work of Bizer, et al.[3]

There has been much existing work in developing the technologies that enable the semantic web. The Resource Description Framework (RDF) is a model of data that provides a way to describe the relationship between resources.[4] RDF allows for the expression of triples in the form of a subject, a predicate, and an object. Once every resource we want to talk about (actors, documents, transactions, policies, etc.) has been associated with a URI, it is possible to use RDF to describe the relationship between these resources (e.g. a subject "transaction", a predicate "compliant with", and an object "Federal Privacy Act"). In addition to providing a way to talk about the relationships between data, we also need a way to describe the hierarchy of objects and how they relate to each other. We do this through the Web Ontology Language (OWL).[5] OWL allows each organization to specify the terms that they are using by way of an ontology, and each organization can also specify the ways entities are related to each other (e.g. a police officer is a sworn law enforcement). In addition, OWL lets us reason between the objects in two different organizations without implicitly assuming that organizations agree on the terminology being used. For example, our system won't assume that a Maryland police officer is interchangeable with a Massachusetts police officer unless that relationship is made explicit.

These notions are particularly important for the applications we're exploring, in that the fundamental problem we're dealing with is data being sent between organizations with different personnel and different information systems. If users, data, and policies can all be referred to in the same language by all organizations in the system, it's not necessary to also warehouse the data in the same place to reason over it. In our system, we are able to assign a URI to each resource we wanted to talk about, so it's possible for each organization in our simulation to keep their data on separate servers. However, systems located at each organization are still able to dereference data on other organizations' systems, and reason over data and personnel from those organizations. This decentralized design does not require a central agency to watch over all transactions to ensure compliance with policy; it's possible for each organization to ensure that the transactions they engage in are compliant with the policies that are relevant to them. In addition, since there is a way for organizations to describe the way they store data and the policies that are relevant to them, it's possible to describe the nuances of each organization and their data in the data itself.

Goals of Accountable Systems

Accountable systems are an alternate way to consider privacy and security in computer systems. Almost all existing systems consider data security to be the problem of safeguarding private information within certain predefined boundaries. However, private data can often be used in certain contexts, but use of that data in other contexts can often be noncompliant with policy. Thus, it is worthwhile to design policies and technology that emphasize accountability rather than impenetrability. Rather than limiting our focus to preventing breaches of private data, we should design systems that are aware of appropriate use and data provenance, so that once a breach occurs, it is easier to determine the source of the problem and deal with the data release after the fact.

Specifically, in this case, we want to use the ideas of accountable systems to give governments increased confidence that they can audit policy-compliant data sharing. For example, if two parties share data about a

person, then a manager, an inspector, or a Court should be able to review why the system concluded that the data sharing event was compliant under the policies governing the transaction. Instead of relying on a "black box" giving a binary assertion about the validity of the transaction, it should be possible to show exactly why the transaction was considered valid under the law. Similarly, if a non-compliant transaction is identified, it should be possible to pinpoint exactly what part of the transaction is questionable, and resolve the matter accordingly.[6]

WORKFLOW OVERVIEW

The implemented system can be queried with hypothetical situations, where a user asks if a document can be sent between two parties. It is assumed that both users have profiles detailing their various affiliations and other relevant information. Such infrastructure already exists in almost every organization, in the form of databases of personnel information. In addition, the document is assumed to be annotated with information that describes the content of the document. The technology to embed machine-readable metadata into document files is already prevalent in commercial document editors. It is also assumed that there is a transcription of the law into computer-readable policy. Such transcriptions can be done by a policy author, and only needs to be done once per policy that needs to be reasoned over.

The user gives URIs (Uniform Resource Identifiers) for each of these components to the system through a web interface. The system then displays a justification of whether or not the hypothetical transaction is valid. For our prototype, using hypotheticals modeled on real world scenarios, a user can see exactly which pieces of the transaction fit together with which clauses of the policy to cause the compliant or non-compliant result.

COMPONENTS OF A DATA TRANSACTION

Rules

The rules governing a data transaction are whatever policies are applicable to the particular data, actors, actions, and context. It is often the case that many different rules from different domains apply simultaneously to a given transaction. For example, if one shares data between two different states, data protection laws from both states need to be applied to the transaction – at a minimum, the law regulating what a sender may release and the law regulating what a receiver may view or store. These policies may use different vocabularies to describe the transaction and may use completely different data sources to reach a sound justification.

Each policy has its own notion of how terms are defined and related. Thus, each policy has to also include an ontology of terms, both so that it can reason about how terms relate to each other (Is a "police officer" a member of "sworn law enforcement"?), and so that two policies' meanings do not become conflated during reasoning.

Data

Historical approaches for applying rules to data focus on categorizing the data in question. For example, in the government arena, rules will be applied broadly across categories such as criminal case investigations or sub-categories describing the type of investigation: e.g., drugs, kidnapping, tax fraud. However, the historical approach falls short of what is present in the law today, and a different approach is necessary. Consider this one segment of a sentence in the Massachusetts law:

> Information shall be provided or made available... only if the individual named in the request or summary has been convicted of a crime punishable by imprisonment for a term of five years or more, or has been convicted of any crime and sentenced to any term of imprisonment, and at the time of the request: is serving a sentence of probation or incarceration, or is under the custody of the parole board... .[7]

This example requires a system implementing the rule to know not only that the general class of data being acted upon falls into the broad class of criminal record, but it also requires the ability to represent information from within the data itself such as: the name of the criminal subject, the

specific statute(s) under which convicted, the length of the sentence imposed, and the current status of the convict.

Entities Described in the Data

This rule requires the system to be able to identify at least three different kinds of people – people who provide information, people who are the recipients of information, and people who are the subjects of the information. Many systems can handle the first two as system users (discussed more below), but have no mechanism to easily communicate the details of a person within target data. Our system uses semantic web techniques to represent these properties. For example, the data subject could be identified as follows:

```
rdf:about=http://dig.csail.mit.edu/2010/
DHS-fusion/MD/CHRI/Guy_Robert_B#rbg
```

which tells the system the URI that identifies Robert B. Guy, the person in the data.

Rule in the Data

Establishing that the person in the data "has been convicted of a crime punishable by imprisonment for a term of five years or more" is done by including a tag to indicate the conviction and the URI for the relevant statute:

```
<mdccl:convicted_pursuant
rdf:resource="http://law.justia.com/
maryland/codes/gps/11-114.html"/>
```

and by including a second tag for the maximum allowable sentence under that statute and the value itself:

```
<mdccl:maximum_allowable_sentence_len
gth>20</
mdccl:maximum_allowable_sentence_lengt
h>
```

Temporal Reasoning

Another determinative fact about the data may require the ability to perform date calculations. Sharing of information is permitted if "at the time of the request: is serving a sentence of probation or incarceration." We represented this as an instance of the subclass which represents custody status:

```
<mdccl:has_custody_status
rdf:resource="mdccl:Parole"/>
```

Also included in the metadata is the sentence imposed:

```
<mdccl:sentence_imposed>5</
mdccl:sentence_imposed>
```

With these two pieces of information, the system has the ability to calculate the end of the sentence based on the date it was imposed and compare that to the "current" date – the date of the request for a data transaction.

Actors

Real rules require the ability to represent details about users of the system at a fine granularity. Again, semantic web technology is well suited to this purpose because it is possible to represent any fact about a user – from the more traditional static values of name, organization, and role, to the discoverable or computable ones such as a person's security keys or the privileges that an actor has within a system. Frequently, rules about data handling are dependent upon what the individual is doing at that moment. For example, the Maryland law allows access to information if and only of it is used in the following two ways:[8]

1. In the performance of its function as a criminal justice agency; or
2. For the purpose of hiring or retaining its own employees and agents.

This sort of information may not be inferable from within a system and may need to be collected as an assertion from the user.

Actions

In order to have meaning, a data usage rule must in some way reference what action is being taken vis-à-vis the target data; it must say an actor can or cannot do something with particular data. These rules refer to actions such as "collect," "retain," "copy," "share," and "delete." Often the action is described using common words, such as "disseminate" or "share," without any definition – for example, without specifying whether these terms apply equally to making data available through push or pull. Because we were modeling the information sharing environment, we focused on sharing rules for the prototype, but could readily represent other actions.

MODELING THE COMPONENT PARTS IN RDF

Rules (AIR)

The rules in our prototype are represented in the AIR (Accountability in RDF) policy language, as described by Kagal, et al.[9] AIR permits the expression of policies as a series of patterns representing criteria to be met for compliance with a particular rule; this works well with legal rules which often are referred to as having "elements", such as the five fair uses of copyright. For the prototype to be accessible for evaluation and validation by a broad array of interested parties (e.g., government executives, policy leaders, lawyers, and the professionals who need to share the information), the sub-rules are coded in the order in which they appear in statute and annotated with their legal citations. This is particularly challenging because law is generated through negotiation and does not generally follow formal logic structures.

We know that some organizations will have the resources and interest to create their own representation of every rule, but that many will opt for a baseline available from a rules library; even in the latter case, there will be law, legal counsel opinion, or policy that is unique to an organization. For this reason, and to demonstrate operation in a decentralized environment, we modeled each organization having a rules library somewhere within the organization's network.

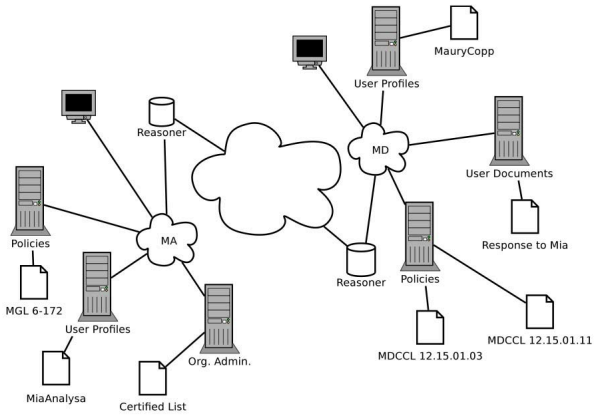

Figure 1: Overview of the Fusion Center System

Actors (FOAF)

The set of attributes in a user profile is normally quite limited in organization databases. We wanted to be able to express essentially anything that might come up in a rule and so chose to adapt FOAF (Friend of a Friend) profiles to represent actors.[10] The FOAF ontology is a relatively short list of attributes, but it is possible to add an unlimited number of additional attributes so long as they are given a URI and, preferably, associated with a definition in a supplemental ontology.

Architecturally, we assumed that each organization would continue to control the user profiles of its employees, members, etc. We did not build, but assumed that each organization would ultimately add, a security layer which determines how much of a profile to reveal to a requesting system. For example, if a foreign organization asks for specific user details in order to reason over them, there should be a system that determines which attributes are revealable and which are private.

Data (PDF/XMP)

Data can be retained in many forms, including email, text documents, databases, and spreadsheets. Since there is already support in commercial software, such as Adobe Acrobat, to annotate documents with RDF, we were not concerned with modeling each structured and unstructured form. For our prototype, we modeled a series of three memos – a request for information about a possible criminal suspect and responses – all in PDF with RDF in an embedded XMP file.

REASONING OVER THE TRANSACTION

Reasoner

A transaction is evaluated against the policy by a forward chaining reasoner, known as cwm.[11] Because of this design choice, the reasoner itself cannot issue calls for more information. Pre-processing must deliver all the necessary data to the reasoner. For example, the prototype automatically identifies to the reasoner the URL for the sender's profile, the proposed recipient's profile, and the target data; it also pre-processes by crawling those files for

references to other policies or ontologies and delivers those URLs to the reasoner as well. In addition, as alluded to earlier, the system searches rules for any assertions that it will need, queries the user, and delivers the result to the reasoner.

TMS

The reasoner has incorporated a Truth Maintenance System,[12] a dependency tracking mechanism. This allows the system to retain the dependencies upon which it relied to form its conclusions. For our prototype, this is extremely useful because it allows users to see the basis for a decision, a function not available from some other policy reasoners. Also, it is an efficient mechanism for storing the necessary information for aggregate reporting, risk modeling, or auditing at a later time.

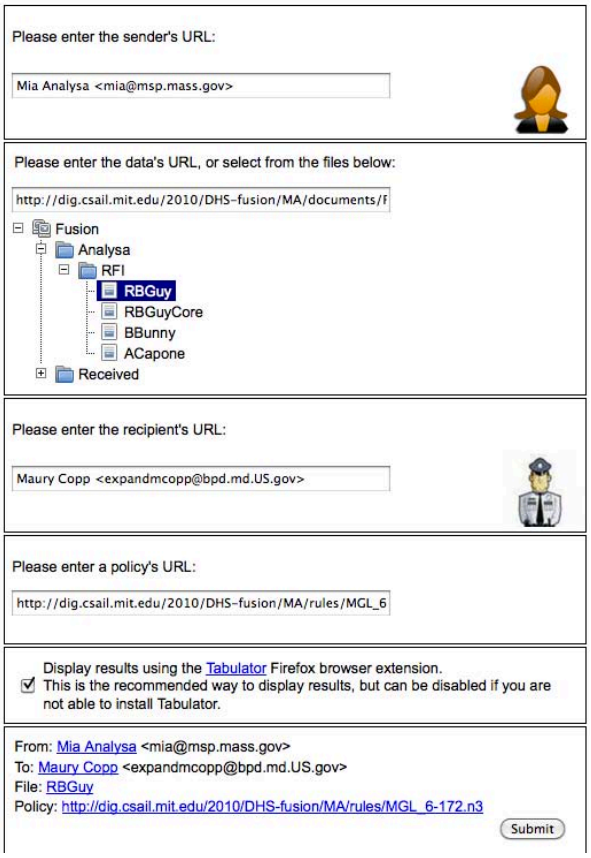

Figure 2: Web Interface for Transactions

VISIBILITY TO USERS

Input: Transaction Simulator

People act on data using many systems and platforms. Rather than separately model transactions in email, various portals, databases, etc., we created a user interface that is intended to provide a view into the middleware, allowing the user to identify the minimal data that would be identified to the accountable system regardless of application or platform – the sender, the target data, and the recipient.

the UI, the sender and receiver are identified by email address, a commonly known identifier, and presumed to be readily linked to the URL for the user profile (FOAF file); the individual's picture and URL are automatically populated on the page. In addition, choosing the data to be sent causes the UI to find and auto-populate the URL for the applicable policy. If necessary to model a variant, the user can override the policy linked to the data with a different policy.

Input: Tabulator Views

Many potential users or evaluators of the technology will not have the skill to read program code. Using a semantic web browser, Tabulator,[13] to view our input code provides an opportunity for those users to glimpse the meaning of the native RDF. Tabulator has multiple viewing panes including a "FOAF View" which makes it possible to see the user profiles in a visualization that looks more like a list of attributes, and a "Table View" which makes it possible to see an AIR policy's if-then-else structure in a nested chart.

Output: Tabulator Views

The accountable system's results can also be viewed in special Tabulator panes. The "Justification" pane first opens to a single sentence that indicates if the proposed transaction is compliant or non-compliant with the policy. Pressing the "Why?" button provides the deep justification provided by the TMS, which shows each belief and facts on which the formation of those beliefs depended.

The "Lawyer" pane provides a shorter form of the analysis by generating a series of

near-grammatical sentences explaining the requirement of the rule, the relevant fact instances that meet or fail to meet that pattern, and the citation for the subsection of the law being applied; the first two are now represented as hyperlinks to the URLs to which they refer.

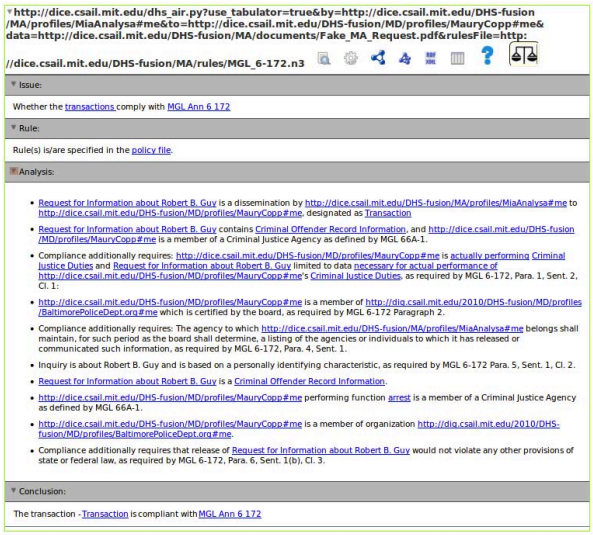

Figure 3: Lawyer View of Justification

RESULTS

Our goal was to model and execute six scenarios through the reasoner; we accomplished that goal and built a system with sufficient capability and flexibility that it is possible to run previously undefined scenarios (mixing and matching the component pieces in unplanned ways) and also achieve correct results.

From the research perspective, this exercise served primarily to confirm expectations. First, it demonstrated some notion of scalability. In our earlier work,[14] we fed to the reasoner only the input necessary to reach a correct conclusion. In this work, we fed the reasoner a significant number of rule patterns and facts that were unnecessary to the conclusion and confirmed, so long as the rules are expressed correctly, that the correct result will be produced – only the appropriate sub-rules will be found to support relevant beliefs, and only the relevant facts will be reported as dependencies. As the "so long as" clause implies, the work showed the importance and necessity of validation, i.e. the ability to determine that the rules have been expressed correctly in their entirety – both the pattern and its relationship to all other patterns (e.g., conditions, exceptions, order). We also proved that "broken" or undefined bits did not necessarily keep the system from reaching a conclusion. For example, if we run a particular scenario under the Massachusetts criminal records release law and the recipient has a malformed tag which was intended to identify him as a member of a criminal justice agency but fails to do so, the system will correctly determine that he is, by a later sub-rule in the policy, entitled to receive such information as any member of the public may receive.

As part of this research, we also demonstrated the prototype to a variety of relevant persons – ranging from Fusion Center analysts to Intelligence Community management, both technical and operational. The reactions were very positive in that such an accountable system could fulfill government obligations to ensure that information sharing is handled in a policy compliant manner and to provide a level of transparency to users.

The most significant resistance received was from an analyst supervisor who perceived this as having the potential to be a management surveillance tool to question the ability of individual analysts to know and comply with all rules; however, even that individual believed that the mechanism would be quite helpful when necessary to apply the rules of another jurisdiction (i.e., not one's own) and for use as a workflow management tool. Conversely, the analysts at a demo the next day were so enthusiastic that they wanted to know if they could build and use the FOAF-based user profiles immediately.

RELATED WORK

Most of the work in this field has focused on building the individual parts of a system, rather than ensuring that all of the pieces are capable of operating together. Our group has published extensively on the importance of a model that is aware that accountability is necessary.[15] There has been work in establishing the importance of the

isomorphism between natural-language law and the machine-readable format,[16] which caused us to consider very carefully whether or not AIR is expressive enough for representing real laws. In addition, there are other reasoning languages similar to AIR such as XACML,[17] and EPAL,[18] which can solve similar problems but without the level of justification granularity presented in this work.

There has also been work in solving similar problems, but in different domains. For example, there exists work that documents the use of semantic web technologies for policy management in the social web.[19] In addition, there is similar work being done for policy enforcement in federated environments,[20] but their work does not seem to have focused on the design of the reasoner.

FUTURE WORK

We learned that the reasoner is relatively fast, for example, processing the potentially more than 100,000 possible pattern match combinations (twenty-seven facts about the sender, twenty-five about the recipient, six about the document, and thirty-five rules) in 10-60 seconds, but that it cannot produce the millisecond response necessary to use the system as a real-time processor for programs that handle millions of transactions daily, such as border applications for customs and screening passengers. However, the Fusion Centers we spoke to indicated that they were producing sufficiently small numbers of analytical reports per day that waiting some seconds for the evaluation would not be prohibitive. We would like to test other reasoning strategies to reduced the time for throughput.

We also are quite interested in coordinating with other test components and systems. Because the United States Constitution establishes state sovereignty, the individual states do not have to follow a federal mandate on the standards for an accountable system and, as a result, we expect that for accountability to be viable there will always be more than one platform, policy language, and tagging scheme in effect. Research is needed to determine the feasibility and strategies for interchange among them.

And, we know that in the physical world among humans, complete information is not always available and yet decisions must be and are made. We would like to learn more about how incomplete information can be effectively handled in an accountable system.

CONCLUSION

Government information sharing is mandated to be performed "[t]o the maximum extent consistent with applicable law." To date, efforts to implement that mandate have been limited by brittle systems that require the system designers to predetermine all likely permissions and then hardwire them. Here, we represented complex policy, reached an array of authoritative sources to implement that complexity, successfully reasoned over the policy, and determined the correct result of compliance/non-compliance. Via this prototype, we have demonstrated the initial feasibility of an accountable system, narrowing the gap between the expectations of law and policy and the ability of technology to fulfill them.

ABOUT THE AUTHORS

K. Krasnow Waterman has had dual careers in technology management and the practice of law. She was the CIO of the first post-9/11 task force created by the president, served as the interim chief operations executive for the reorganization of FBI intelligence infrastructure, and represented the Department of Homeland Security in high-level negotiations to set the requirements for interoperability of federal data systems. She now divides her time between managing large-scale operations infrastructure and researching new web technology at MIT's Computer Science lab. She may be contacted at kkw@MIT.EDU.

Samuel Wang is a second year PhD student at MIT, working as a member of the Decentralized Information Group in the Computer Science and Artificial Intelligence Laboratory. His primary interests relate to web technologies and natural language processing. Before attending MIT, he completed his undergraduate work at Carnegie Mellon University in Computer Science and Mathematics. In addition, he worked at Google as a software engineer, primarily on various aspects of web search.

[1] White House, "Strengthening the Sharing of Terrorism Information to Protect Americans," Executive Order 13356 (Washington, DC: The White House, 2004) and *Intelligence Reform and Terrorism Prevention Act,* Public Law 108-458, 108th Congress, 1st sess. (December 17, 2004).

[2] Code of Maryland Regulations (COMAR), Title 12, Section 15.01.11 (2009), http://www.dsd.state.md.us/comar/comarhtml/12/12.15.01.11.htm.

[3] C. Bizer, T. Heath, and T. Berners-Lee, "Linked Data – The Story So Far," *International Journal on Semantic Web and Information Systems* (2009).

[4] O. Lassila and R. Swick, "Resource Description Framework (RDF) Model and Syntax," W3C - World Wide Web Consotrium (1999),http://www.w3.org/TR/1999/REC-rdf-syntax-19990222/.

[5] D. McGuinness, et al., "OWL Web Ontology Language Overview," W3C Recommendation (February 10, 2004), http://www.w3.org/TR/owl-features.

[6] A more comprehensive treatment of accountable systems can be found in the paper by D. Weitzner et al., "Information Accountability," *Communications of the ACM* 51, no. 6 (2008): 82-87.

[7] Massachusetts General Laws, "Administration of the Government," Part 1, Title II, Chapter 6, Section 172 (2009), http://www.malegislature.gov/Laws/GeneralLaws/Part1/TitleII/Chapter6/Section172.

[8] COMAR, Title 12, 2009

[9] L. Kagal, C. Hansen, and D. Weitzner, "Using Dependency Tracking to Provide Explanations for Policy Management," *Proceeding of the IEEE Workshop on Policies for Distributed Systems and Networks* (Palisades, NY, June 2-4, 2008), http://doe.ieeecomputersociety.org/10.1109.POLICY.2008.51 .

[10] D. Brickley and L. Miller, "FOAF Vocabulary Specification 0.91," Namespace Document (FOAF Project: November 2007), http://xmlns.com/foaf/0.1.

[11] T. Berners-Lee, et al., "Cwm: A General Purpose Data Processor for the Semantic Web," Project Web Site, W3C (2006),http://www.w3.org/2000/10/swap/doc/cwm.html.

[12] J. Doyle, "A Truth Maintenance System," *Artificial Intelligence* 12, no. 3 (1979): 231-272.

[13] T. Berners-Lee, et al., "Tabulator: Exploring and Analyzing Linked Dataon the Semantic Web," *Proceedings of the 3rd International Semantic Web User Interaction Workshop* (Athens, Georgia: November 6, 2006).

[14] D. Weitzner, et al., "Transparent Accountable Data Mining: New Strategies for Privacy Protection," in *AAAI Spring Symposium on the Semantic Web Meets eGovernment* Technical Report SS-06-06 (Menlo Park, CA: AAAI Press, 2006), http://www.aaai.org/Papers/Symposia/Spring/2006/SS-06-06/SS06-06-025.pdf.

[15] Ibid.

[16] T. Bench-Capon and F. Coenen, "Isomorphism and Legal Knowledge Based Systems," *Artificial Intelligence and Law* 1, no. 1 (1992): 65-86.

[17] S. Godik, et al., "Extensible Access Control Markup Language (xacml) Version 1.0," OASIS Standard (OASIS Open, 2003), www.oasis-open.org/committees/download.php/944.

[18] P. Ashley, et al., "Enterprise Privacy Authorization Language (EPAL 1.2)," Submission to W3C (2003), www.w3.org/Submission/2003/SUBM-EPAL-20031110/.

[19] J. Zeiss, et al., "A Semantic Policy Management Environment for End-Users," *Proceedings of International Conference on Semantic Systems* (2008), http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.139.3129.

[20] V. Hu, S. Quirolgico, and K. Scarfone, "Access Control Policy Composition for Resource Fedeartion Networks Using Semantic Web and Resource Description Framework (RDF)," *Proceedings of the 2008 International Computer Symposium* (Taiwan, November 13-15, 2008).